

SBIR Technical Questions for NIH

November 5, 2003

1. The *Explanation* element in 1.2.1.2.3.2 and 1.2.1.2.4.2 is related to Debarment and Drug Free question. We had previously pointed out that these question were not listed and the phs398 crosswalk, and were therefore not important to the validation of the proposal. If this is the case then why is the business rule validating these elements data?

Answer: The viewpoint of the validations is, regardless of how these answers arrive, to enforce consistency if they do. It would be better if we did not validate these particular components at all, but that will have to be taken up post-pilot. The workaround is simple though, simply always provide “true” as the response to these two assurance questions, and an explanation will not be required. These answers do not form any part of the PHS 398 application image nor will they be reviewed when evaluating the proposal.

2. Rule Level 1.2.1.4 validates the KeyPerson and states there can only be one PI or one PD listed. Rule Level 1.2.1.4.13 validates the KeyPerson as either a “PI” or “KP”. Given this scenario, doesn’t this then mean the 1.2.1.4 can only be PI?

Answer: Yes

3. The organization Contact name discussed in rule 1.2.1.5.13.11.1, and states the combinations of all child elements must be 30 characters or less. Here is where it gets a little confusing to me. First I am not sure the form itself imposes this restriction on the applicant. My second concern here is that we are required to have the components of a name separated into there own fields, and then we put a restriction on the group.

Answer: First concern: if the comment regarding “the form itself” is referring to the paper PHS 398 form, then no, it does not place any such restriction on the applicant. For that matter, very few length restrictions are shared with the applicant in the paper process. However, in the electronic process we are going to need to guard our storage constraints more closely; we do not have the same opportunity in this scenario, to apply human effort to sanitizing/editing information before it gets to the NIH database.

Second concern: I think the main point of confusion is, why do we collect the name as separate fields, but combine for storage. The answer is consistency—that we would like to standardize on one method of name collection in the schema, using the last, first, middle, etc., construct. This, in fact, is the direction we will have to take when integrating with Grants.gov, so we may as well take this approach now. At some point in the future, the issue should be resolved by making the storage of the name in the NIH database match the manner in which it is collected (which would have avoided the confusion now). This was not undertaken for the pilot due to the existing data modeling for this contact name information (all one field) and our desire to minimize the impact on existing, downstream NIH processes (which currently depend on the name being stored as one long field and not broken down).

The length of the individual sections therefore is not important, but the length of all the sections (added-up) is important.

4. Email element is described in rule version 1.2.1.5.13.112.1 is stated as having a max size of 80 characters, but the rules states that, “Must contain a ‘@’, with at least 1 and at most 64 characters preceding and following the ‘@’, Control characters (ASCII 0 through 31 and 127), spaces and the special characters < > () [] \ , ; : are not valid.” Does this mean [1-64]&[1-64] or something else? Which is correct, the 80 or the 64/128?

Answer: Both restrictions apply. The total max characters is 80 because this is our database column limit, and there can be at most 64 characters before or after the @ as long as the total is not greater than 80 (e.g., [64]@[10] is valid, but [64]@[16] is not). The 64-char limit is per the ANSI standard.

5. Rule Level 1.2.1.3.3.6.3 states that a ProjectRole must have one of the listed codes. However, I have a proposal with “PI” and “PDA” as ProjectRoles and they are not validating. Must be a valid role (in uppercase), as defined below.

PI	Principal Investigator
AWD	Awardee
PD	Program Director
FEL	Fellow
AO	Authorizing Official
SUP	Supportee
TA	Trainee Appointment
LT	Lab Technician
RA	Research Assistant
AA	Administrative Assistant
SO	Signing Official
SS	Staff Scientist
KP	Key Personnel
CPI	Co Principal Investigator
NKP	Non key personnel
CSU	Consultant
PDC	Post-Doctoral
PDA	Post-Doctoral Associate
OP	Other Professional
GS	Graduate Student
UGS	Undergraduate Student
SC	Secretarial/Clerical
OTH	Other

Answer: This proposal will fail for a bigger reason than the ProjectRole business rule. The main problem that we see is that for a submission during the pilot, component 1.2.1.3.3.6.3 should not be submitted at all. Component 1.2.1.3.3.6.3 is the ProjectRole for a person, for whom SalariesAndWages information is being submitted. But SalariesAndWages is a component submitted with a detailed, non-modular budget (page 4), and therefore violates the parameters of the pilot. The answer to this problem is (1) verify that the budget meets the definition of modular, and (2) if so, submit ONLY the XML components which are applicable to the modular budget page.

6. We are concerned about error messages we are receiving in the XML from the NIH. We plan to perform the same validation as is being performed by NIH so we can alert our customer immediately to problems before she attempts to submit a grant electronically. However, it is inevitable that some will slip through and NIH will detect errors and send them back to us. Our concern is the translation of these error messages into messages that our customer can readily understand.

To solve this problem, we'd like to get a catalog of NIH error messages and the context when each error message is generated. Do you have such a catalog or planning to create one? The error messages we've seen so far include res:id, res:errorCode, res_location and res:codeContext attributes. Can you (have you already) document what these values mean for each error? The res:location has an interesting name, hopefully this will eventually pinpoint where in the grant application the error is located.

Here's an example: I'm looking at a 398 Modular face page. In box 2, the user can check "**Yes**" and enter a number and a Title. If the number is invalid, the error returned by NIH is "Funding opportunity number is invalid." If you look carefully at box 2 on a 398 face page, you'll see that neither the word "opportunity" nor the word "funding" appear there. If our customer can get this error, then we need to translate it so he understands what form field he has entered incorrectly.

Answer: #6 is asking if we're putting together an error catalog so that error messages can be better presented to the applicant. I don't believe so. I think our strategy is, for the pilot, leave the error messages alone essentially, since we're now past the deadline for all submissions (the value of continuing to tinker with them is drawing to a close). The long-term strategy, to be rolled in prior to the next pilot, is to produce much more descriptive error messages to replace the ones we currently have. This should diminish the need to produce an error catalog, although it still seems as though it would be a valuable piece of documentation to produce at *some* point.

7. Do you have a schedule of what's going to be happening over the next 4–8 months? We are interested in the plans for future pilot programs, that is, start date, end date, and if a new type of form will be accepted during the pilot, assuming that we are going to continue with these pilot programs. Beyond the next pilot, the dates are less important, and we are primarily interested in which new form(s) each subsequent pilot will accept.

Answer:

Timeframe	Action
November-mid January	NIH to correct the most critical issues encountered in the exchange software, during the pilot period.
Mid-January to January 31	Testing with SBIRs, using next round of Type-1 (new) proposals to be submitted, for February receipt date. Scope of pilot is not expanding. We are still looking for a limited number of applications to be submitted, following the same R01/modular budget requirements that we've had for this first pilot.
February	Submit the Type-1 applications received in January, error-free, through the eRA exchange in production

Timeframe	Action
February	Continue to test with the SBIRs, using the next round of Type-2 (renewal) applications to be submitted, for the March receipt date. Still interested in small number of applications, R01 type and modular budgets
March	Submit the Type-2 applications received in February, error-free, through the eRA exchange in production

8. I've just been in touch with University of Michigan. They are frustrated both by the current and proposed lists of valid entries and they point to the published guidelines for the 398 which state: "Under role on the project, indicate how the individual will function with regard to the proposed project."

As they suggest, roles on projects will vary greatly. They go on to say: "I hope that when the CWG was discussing this section, we did not explicitly (or tacitly) endorse the notion of the predetermined pick list. This would represent a great annoyance to the PIs."

I think their concerns are entirely valid and I wonder if the proposed "Other" classification under <ProjectRole>, followed by the optional <ProjectRoleDescription> is the mechanism for handling the broad range of roles in the real world.

All of this goes on top of my earlier concerns about matching.

Clarification: The instructions are not the problem but they imply greater flexibility than currently provided by either the Commons or GrantSlam (which we built to the Commons spec). The latest spec, provided in CGAP validation.doc, may be sufficiently broad, due to its inclusion of "Other."

However, Michigan suggests that Other may be used quite frequently because PIs are not happy with limited choices. One of the proposals we'll be submitting later this week has these entries under Role on Project:

"Principal Investigator, Supervision of Research Purification of lectins"—and— "Research Investigator, Use of *Marasmius oreades* lectin in a model for glomerular microangiopathic injury in mice."

As mentioned in my earlier message, the discrepancies between what constitutes a valid entry on the Commons and in GrantSlam on the one hand vs. the validation document on the other, have me concerned about what to code for this element, and what to submit.

Answer: The proposed way of expanding the list is not really the answer. The reference here is to the "role on project/project role description" pair, which qualifies the salaries and wages component. This component applies only to detailed budgets so it is not a solution that is general enough to satisfy all situations, especially during the pilot phase when *all* submitted budgets are *modular*, not detailed.

It is understood that the "role on project" column for key personnel must ultimately provide a richer set of choices, and an ability to provide roles that fall beyond the limited scope of the picklist. This ties in very closely with issues being addressed at this time by the IAEGC Research and Related data set committee and we will align ourselves with whatever final decisions are reached by that committee. At this time, the resolution of this issue within the R&R dataset is still uncertain, so

for the duration of the current pilot, the role of KP will be deemed sufficient to identify the role of key personnel other than the PI. While this designation does not provide the level of detail normally expressed in the Key Personnel section, the applicant has NIH's assurance that this limitation will in no way present a disadvantage to the electronic applicant.

The Personnel section of the Budget Justification may be used to express the role of each Key Person in greater detail, if so desired.

9. We are submitting a proposal that contains no cover letter. As you are probably aware, the instructions for completing the PHS 398 have this to say about a Cover Letter:

“If the principal investigator is making a request for assignment to a particular awarding component or initial review group, include a cover letter.”

That first word, “IF,” implies the letter is optional. Why is it not optional in the schema?

Answer: Yes, I agree that this should be optional in the schema. As we've gone through this process, we have discovered that the optional/mandatory facets for some of the attachments do require some re-examination post-pilot.

Most were defined required by convention at the beginning, since making a component optional needed some further consideration, on a case-by-case basis. Needless to say, I've not had a chance to revisit these original defaults, in the midst of getting all the other pilot issues resolved. The option of attachments is an important issue to get resolved, but at this stage it is safest to defer any changes until after the current pilot is completed.

You probably will see some schema changes in the area of cover letters (as well as other attachments) in the next pilot phase, which will occur sometime after November. For the current phase, if there is no cover letter, the workaround is easy. Since the file identifier is a string, and the string has no minLen facet defined, you can simply supply an empty tag to satisfy schema validation. There is no subsequent validation on the back-end which will go any further to assert the presence of a cover letter, so you should be home free at that point.

The following question was answered in an email to all participants by Gerald prior to the meeting. I thought it should be included in the Q&A but probably does not need to be reviewed at the meeting.

10. The competing continuation qualifiers component is optional but, if present, the indicated subcomponents must also be present. It is made up of the application type code, activity code, administering organization id, serial number, support year, and suffix code. The Grant number on the form is 1 field that contains the entire grant number; however the XML is being required to break this number into components. Below is the table that states what the length of each component is. My question is, what are the rules for breaking apart this number? Can I use the lengths defined below as an absolute guide to break apart the number, and are these components listed in order?

1.2.2.2.1	ApplicationTypeCode	String	1	No
1.2.2.2.2	ActivityCode	String	3	No

1.2.2.2.3	AdministeringOrganizationID	String	2	No
1.2.2.2.4	SerialNumber	Long	6	No
1.2.2.2.5	SupportYear	Integer	2	No
1.2.2.2.6	SuffixCode	String	4	No

Answer: We can take a shot at breaking the number up by lengths, but this could be dicey in a couple spots. Perhaps a better parsing algorithm could be devised though, based on the general format of a grant number (as follows):

1R01MH012345-01A1S1

Byte	Description
Byte 1:	ApplicationTypeCode. Always numeric, in range 1..9
Bytes 2–4:	ActivityCode. Always 3-character alphanumeric, starting with an ALPHA character, ending in a DIGIT. Second character can be either ALPHA or DIGIT.
Bytes 5–6:	AdministeringOrganizationID. Always 2 ALPHA characters, NO DIGITS.
Bytes 7–12:	SerialNumber. Always numeric. Generally, correspondence with the grantee will use leading zeroes so applicants are likely to follow suit. But it is possible that the applicant will only express the significant digits (i.e., 12345 instead of 012345).
Byte 13:	Dash. Not stored in our database but ALWAYS appears on screens, reports, and correspondence. This should be counted on when performing data collection.
Bytes 14–15:	SupportYear. Always numeric, greater than zero. Generally, this should use a leading zero if less than 10, but again, there is the possibility that an applicant may give only one digit if <= 9.
Bytes 16–19:	SuffixCode: Optional. Not all grants have suffixes. The possible lengths are 0, 2, or 4, and always alphanumeric. If a suffix is provided, the first byte must always be a character, and the second byte a digit, range 1..9. This pattern repeats for the third and fourth bytes, if those are provided.